ACC gene

The present invention relates to a gene useful in a process to increase the microbial production of carotenoids.

The carotenoid astaxanthin is distributed in a wide variety of organisms such as animals, algae and microorganisms. It has a strong antioxidation property against reactive oxygen species. Astaxanthin is used as a coloring reagent, especially in the industry of farmed fish, such as salmon, because astaxanthin imparts distinctive orange-red coloration to the animals and contributes to consumer appeal in the marketplace.

One of the first steps in the carotenogenic pathway of, e.g. *Phaffia rhodozyma*, is the condensation of two molecules of acetyl-CoA. Acetyl-CoA is also the substrate for acetyl-CoA carboxylase, one of the enzymes involved in fatty acid biosynthesis.

In one aspect, the present invention provides a novel DNA fragment comprising a gene encoding the enzyme acetyl-CoA carboxylase.

More particularly, the present invention provides a DNA containing regulatory regions, such as promoter and terminator, as well as the open reading frame of acetyl-CoA carboxylase gene.

The present invention provides a DNA fragment encoding acetyl-CoA carboxylase in *P. rhodozyma*. The said DNA means a cDNA which contains only open reading frame flanked between the short fragments in its 5'- and 3'- untranslated region, and a genomic DNA which also contains its regulatory sequences such as its promoter and terminator which are necessary for the expression of the acetyl-CoA carboxylase gene in *P. rhodozyma*.

Accordingly, the present invention relates to a polynucleotide comprising a nucleic acid molecule selected from the group consisting of:

(a) nucleic acid molecules encoding at least the mature form of the polypeptide depicted in SEQ ID NO:3;

(b) nucleic acid molecules comprising the coding sequence as depicted in SEQ ID NO:2;

(c) nucleic acid molecules whose nucleotide sequence is degenerate as a result of the genetic code to a nucleotide sequence of (a) or (b);

(d) nucleic acid molecules encoding a polypeptide derived from the polypeptide encoded by a polynucleotide of (a) to (c) by way of substitution, deletion and/or addition of one or several amino acids of the amino acid sequence of the polypeptide encoded by a polynucleotide of (a) to (c);

(e) nucleic acid molecules encoding a polypeptide derived from the polypeptide whose sequence has an identity of 56.3 % or more to the amino acid sequence of the polypeptide encoded by a nucleic acid molecule of (a) or (b);

(f) nucleic acid molecules comprising a fragment or a epitope-bearing portion of a polypeptide encoded by a nucleic acid molecule of any one of (a) to (e) and having acetyl-CoA carboxylase activity;

(g) nucleic acid molecules comprising a polynucleotide having a sequence of a nucleic acid molecule amplified from *Phaffia* or *Xanthophylomyces* nucleic acid library using the primers depicted in SEQ ID NO:4, 5, and 6;

(h) nucleic acid molecules encoding a polypeptide having acetyl-CoA carboxylase activity, wherein said polypeptide is a fragment of a polypeptide encoded by any one of (a) to (g);

(i) nucleic acid molecules comprising at least 15 nucleotides of a polynucleotide of any one of (a) to (d);

(j) nucleic acid molecules encoding a polypeptide having acetyl-CoA carboxylase activity, wherein said polypeptide is recognized by antibodies that have been raised against a polypeptide encoded by a nucleic acid molecule of any one of (a) to (h);

(k) nucleic acid molecules obtainable by screening an appropriate library under stringent conditions with a probe having the sequence of the nucleic acid molecule of any one of (a) to (j), and encoding a polypeptide having an acetyl-CoA carboxylase activity;

(l) nucleic acid molecules whose complementary strand hybridizes under stringent conditions with a nucleic acid molecule of any one of (a) to (k), and encoding a polypeptide having acetyl-CoA carboxylase activity.

The terms "gene(s)", "polynucleotide", "nucleic acid sequence", "nucleotide sequence", "DNA sequence" or "nucleic acid molecule(s)" as used herein refers to a polymeric form of

nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule.

Thus, this term includes double- and single- stranded DNA, and RNA. It also includes known types of modifications, for example, methylation, "caps" substitution of one or more of the naturally occurring nucleotides with an analog. Preferably, the DNA sequence of the invention comprises a coding sequence encoding the above-defined polypeptide.

A "coding sequence" is a nucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, recombinant nucleotide sequences or genomic DNA, while introns may be present as well under certain circumstances. SEQ ID:1 depicts the genomic DNA in which the intron sequence is inserted in the coding sequence for acetyl-CoA carboxylase gene from *P. rhodozyma*.

In general, the gene consists of several parts which have different functions from each other. In eukaryotes, genes which encode a corresponding protein, are transcribed to premature messenger RNA (pre-mRNA) differing from the genes for ribosomal RNA (rRNA), small nuclear RNA (snRNA) and transfer RNA (tRNA). Although RNA polymerase II (PolII) plays a central role in this transcription event, PolII can not solely start transcription without *cis* element covering an upstream region containing a promoter and an upstream activation sequence (UAS), and a *trans*-acting protein factor. At first, a transcription initiation complex which consists of several basic protein components recognize the promoter sequence in the 5'-adjacent region of the gene to be expressed. In this event, some additional participants are required in the case of the gene which is expressed under some specific regulation, such as a heat shock response, or adaptation to a nutrition starvation, and so on. In such a case, a UAS is required to exist in the 5'-untranslated upstream region around the promoter sequence, and some positive or negative regulator proteins recognize and bind to the UAS. The strength of the binding of transcription initiation complex to the promoter sequence is affected by such a binding of the *trans*-acting factor around the promoter, and this enables the regulation of transcription activity.

After the activation of a transcription initiation complex by the phosphorylation, a transcription initiation complex initiates transcription from the transcription start site. Some parts of the transcription initiation complex are detached as an elongation complex from the promoter region to the 3' direction of the gene (this step is called as a promoter

clearance event) and the elongation complex continues the transcription until it reaches to a termination sequence that is located in the 3'-adjacent downstream region of the gene. Pre-mRNA thus generated is modified in nucleus by the addition of cap structure at the cap site which almost corresponds to the transcription start site, and by the addition of polyA stretches at the polyA signal which is located at the 3'-adjacent downstream region. Next, intron structures are removed from the coding region and exon parts are combined to yield an open reading frame whose sequence corresponds to the primary amino acid sequence of a corresponding protein. This modification in which a mature mRNA is generated is necessary for a stable gene expression. cDNA in general terms corresponds to the DNA sequence which is reverse-transcribed from this mature mRNA sequence. It can be synthesized by the reverse transcriptase derived from viral species by using a mature mRNA as a template, experimentally.

To express a gene which was derived from eukaryote, a procedure in which cDNA is cloned into an expression vector for *E. coli* is often used. This results from the fact that a specificity of intron structure varies among the organisms and an inability to recognize the intron sequence from other species. In fact, prokaryote has no intron structure in its own genetic background. Even in yeast, the genetic background is different between *Ascomycetes* to which *Saccharomyces cerevisiae* belongs and *Basidiomycetes* to which *P. rhodozyma* belongs, e.g. the intron structure of the actin gene from *P. rhodozyma* cannot be recognized nor spliced by the ascomycetous yeast, *S. cerevisiae.*

Intron structures of some kinds of the genes appear to be involved in the regulation of the expression of their genes. It might be important to use a genomic fragment which has its introns in a case of self-cloning of the gene of a interest whose intron structure involves such a regulation of its own gene expression.

To apply a genetic engineering method for a strain improvement study, it is necessary to study its genetic mechanism in the event such as transcription and translation. It is important to determine a genetic sequence such as its UAS, promoter, intron structure and terminator to study the genetic mechanism.

According to this invention, the gene encoding the acetyl-CoA carboxylase (*ACC*) gene from *P. rhodozyma* including its 5'- and 3'-adjacent regions as well as its intron structure was determined.

The invention further encompasses polynucleotides that differ from one of the nucleotide sequences shown in SEQ ID NO:2 (and portions thereof) due to degeneracy of the genetic

code and also encode an acetyl-CoA carboxylase as that encoded by the nucleotide sequences shown in SEQ ID NO:2. Further the polynucleotide of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:3. In a still further embodiment, the polynucleotide of the invention encodes a full length *P. rhodozyma* protein which is substantially homologous to an amino acid sequence of SEQ ID NO:3.

In addition, it will be appreciated by those skilled in the art that DNA sequence polymorphism that lead to changes in the amino acid sequences may exist within a population (e.g., the *P. rhodozyma* population). Such genetic polymorphism in the acetyl-CoA carboxylase gene may exist among individuals within a population due to natural variation.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an acetyl-CoA carboxylase, preferably an acetyl-CoA carboxylase from *P. rhodozyma.*

Such natural variations can typically result in 1-5 % variance in the nucleotide sequence of the acetyl-CoA carboxylase gene. Any and all such nucleotide variations and resulting amino acid polymorphism in acetyl-CoA carboxylase that are the result of natural variation and that do not alter the functional activity of acetyl-CoA carboxylase are intended to be within the scope of the invention.

Polynucleotides corresponding to natural variants and non-*P. rhodozyma* homologues of the acetyl-CoA carboxylase cDNA of the invention can be isolated based on their homology to *P. rhodozyma* acetyl-CoA carboxylase polynucleotides disclosed herein using the polynucleotide of the invention, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, a polynucleotide of the invention is at least 15 nucleotides in length. Preferably it hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of the polynucleotide of the present invention, e.g. SEQ ID NO:2. In other embodiments, the nucleic acid is at least 20, 30, 50, 100, 250 or more nucleotides in length. The term "hybridizes under stringent conditions" is defined above and is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% identical to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65% or 70%, more preferably at least about 75% or 80%, and even more preferably at least about 85%, 90% or 95% or more identical to each other typically remain hybridized to each other. Preferably, polynucleotide of the invention that hybridizes under stringent

conditions to a sequence of SEQ ID NO:2 corresponds to a naturally occurring nucleic acid molecule.

In the present invention, the polynucleotide sequence includes SEQ ID NO:2 and fragments thereof having polynucleotide sequences which hybridize to SEQ ID NO:2 under stringent conditions which are sufficient to identify specific binding to SEQ ID NO:2. For example, any combination of the following hybridization and wash conditions may be used to achieve the required specific binding:

High Stringent Hybridization: 6X SSC, 0.5% SDS, 100 µg/ml denatured salmon sperm DNA, 50% formamide, incubate overnight with gentle rocking at 42°C.

High Stringent Wash: 1 wash in 2X SSC, 0.5% SDS at room temperature for 15 minutes, followed by another wash in 0.1X SSC, 0.5% SDS at room temperature for 15 minutes.

Low Stringent Hybridization: 6X SSC, 0.5% SDS, 100 µg /ml denatured salmon sperm DNA, 50% formamide, incubate overnight with gentle rocking at 37°C.

Low Stringent Wash: 1 wash in 0.1X SSC, 0.5% SDS at room temperature for 15 minutes.

Moderately stringent conditions may be obtained by varying the temperature at which the hybridization reaction occurs and/or the wash conditions as set forth above. In the present invention, it is preferred to use high stringent hybridization and wash conditions to define the antisense activity against acetyl-CoA carboxylase gene from *P. rhodozyma.*

The term "homology" means that the respective nucleic acid molecules or encoded proteins are functionally and/or structurally equivalent. The nucleic acid molecules that are homologous to the nucleic acid molecules described above and that are derivatives of said nucleic acid molecules are, for example, variations of said nucleic acid molecules which represent modifications having the same biological function, in particular encoding proteins with the same or substantially the same biological function. They may be naturally occuring variations, such as sequences from other plant varieties or species, or mutations. These mutations may occur naturally or may be obtained by mutagenesis techniques. The allelic variations may be naturally occurring allelic variants as well as synthetically produced or genetically engineered variants. Structural equivalents can, for example, be identified by testing the binding of said polypeptides to antibodies. Structural equivalents have similar immunological characteristics, e.g. comprise similar epitopes.

As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). Preferably, the polynucleotide encodes a natural *P. rhodozyma* acetyl-CoA carboxylase.

In addition to naturally-occurring variants of the acetyl-CoA carboxylase sequence that may exist in the population; the skilled artisan will further appreciate that changes can be introduced by mutation into a nucleotide sequence of the polynucleotide encoding acetyl-CoA carboxylase, thereby leading to changes in the amino acid sequence of the encoded acetyl-CoA carboxylase, without altering the functional ability of the acetyl-CoA carboxylase. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of the polynucleotide encoding acetyl-CoA carboxylase, e.g. SEQ ID NO:2. A "non- essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the acetyl-CoA carboxylase without altering the activity of said acetyl-CoA carboxylase, whereas an "essential" amino acid residue is required for acetyl-CoA carboxylase activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having acetyl-CoA carboxylase activity) may not be essential for activity and thus are likely to be amenable to alteration without altering acetyl-CoA carboxylase activity.

Accordingly, the invention relates to polynucleotides encoding acetyl-CoA carboxylase that contain changes in amino acid residues that are not essential for acetyl-CoA carboxylase activity. Such acetyl-CoA carboxylase differs in amino acid sequence from a sequence contained in SEQ ID NO:3 yet retain the acetyl-CoA carboxylase activity described herein. The polynucleotide can comprise a nucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least about 60% identical to an amino acid sequence of SEQ ID NO:3 and has acetyl-CoA carboxylase activity. Preferably, the protein encoded by the nucleic acid molecule is at least about 60-65% identical to the sequence in SEQ ID NO:3, more preferably at least about 60-70% identical to one of the sequences in SEQ ID NO:3, even more preferably at least about 70-80%, 80- 90%, 90-95% homologous to the sequence in SEQ ID NO:3, and most preferably at least about 96%, 97%, 98%, or 99% identical to the sequence in SEQ ID NO:3.

To determine the percent homology of two amino acid sequences (e.g., one of the sequence of SEQ ID NO:3 and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of SEQ ID NO:2 or 3) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected), then the molecules are homologous at that position (i.e., as used herein amino

acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = numbers of identical positions/total numbers of positions x 100). The homology can be determined by computer programs as Blast 2.0 [Altschul, Nuc. Acid. Res., 25:3389-3402 (1997)]. In this invention, GENETYX-SV/RC software (Software Development Co., Ltd., Tokyo, Japan) is used by using its default algorithm as such homology analysis software. This software uses the Lipman-Pearson method for its analytic algorithm.

A nucleic acid molecule encoding an acetyl-CoA carboxylase homologous to a protein with an amino acid sequence of SEQ ID NO:3 can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of the polynucleotide of the present invention, in particular of SEQ ID NO:2 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into the sequences of, e.g., SEQ ID NO:2 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an acetyl-CoA carboxylase is preferably replaced with another amino acid residue from the same family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an acetyl-CoA carboxylase coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an acetyl-CoA carboxylase activity described herein to identify mutants that retain acetyl-CoA carboxylase activity. Following mutagenesis of one of the sequences of SEQ ID NO:2, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein.

A polynucleotide of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of SEQ ID NO:2, or a portion thereof, can be isolated using standard molecular

biology techniques and the sequence information provided herein. For example, acetyl-CoA carboxylase cDNA can be isolated from a library using all or portion of one of the sequences of the polynucleotide of the present invention as a hybridization probe and standard hybridization techniques. Moreover, a polynucleotide encompassing all or a portion of one of the sequences of the polynucleotide of the present invention can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of polynucleotide of the present invention can be isolated by the polymerase chain reaction using oligonucleotide primers, e.g. of SEQ ID NO:4, 5, or 6, designed based upon this same sequence of polynucleotide of the present invention. For example, mRNA can be isolated from cells, e.g. *Phaffia* (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. and cDNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase or AMV reverse transcriptase available from Promega (Madison, USA)). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in SEQ ID NO:2. A polynucleotide of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The polynucleotide so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an acetyl-CoA carboxylase nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

The terms "fragment", "fragment of a sequence" or "part of a sequence" means a truncated sequence of the original sequence referred to. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; the minimum size being a sequence of sufficient size to provide a sequence with at least a comparable function and/or activity of the original sequence referred to, while the maximum size is not critical. In some applications, the maximum size usually is not substantially greater than that required to provide the desired activity and/or function(s) of the original sequence.

Typically, the truncated amino acid sequence will range from about 5 to about 60 amino acids in length. More typically, however, the sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select sequences of at least about 10, 12 or 15 amino acids, up to maximum of about 20 or 25 amino acids.

The term "epitope" relates to specific immunoreactive sites within an antigen, also known as antigenic determinants. These epitopes can be a linear array of monomers in a polymeric composition - such as amino acids in a protein - or consist of or comprise a more complex secondary or tertiary structure. Those of skill will recognize that all immunogens (i. e., substances capable of eliciting an immune response) are antigens; however, some antigen, such as haptens, are not immunogens but may be made immunogenic by coupling to a carrier molecule. The term "antigen" includes references to a substance to which an antibody can be generated and/or to which the antibody is specifically immunoreactive.

The term "one or several amino acids" relates to at least one amino acid but not more than that number of amino acids which would result in a homology of below 60% identity. Preferably, the identity is more than 70% or 80%, more preferred are 85%, 90% or 95%, even more preferred are 96%, 97%, 98%, or 99% identity.

The term "acetyl-CoA carboxylase" or "acetyl-CoA carboxylase activity" relates to enzymatic activities of a polypeptide as described below or which can be determined in enzyme assay method. Furthermore, polypeptides that are inactive in an assay herein but are recognized by an antibody specifically binding to acetyl-CoA carboxylase, i.e., having one or more acetyl-CoA carboxylase epitopes, are also comprised under the term "acetyl-CoA carboxylase". In these cases activity refers to their immunological activity.

The terms "polynucleotide" and "nucleic acid molecule" also relate to "isolated" polynucleotides or nucleic acids molecules. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived.

For example, in various embodiments, the PNO polynucleotide can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g., a *Phaffia* cell). Moreover, the polynucleotides of the present invention, in particular an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

Preferably, the polypeptide of the invention comprises one of the nucleotide sequences shown in SEQ ID NO:2. The sequence of SEQ ID NO:2 corresponds to the *P. rhodozyma* acetyl-CoA carboxylase cDNAs of the invention.

Further, the polynucleotide of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences of above mentioned polynucleotides or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in SEQ ID NO:2 is one which is sufficiently complementary to one of the nucleotide sequences shown in SEQ ID NO:2 such that it can hybridize to one of the nucleotide sequences shown in SEQ ID NO:2, thereby forming a stable duplex.

The polynucleotide of the invention comprises a nucleotide sequence which is at least about 60%, preferably at least about 65-70%, more preferably at least about 70-80%, 80-90%, or 90-95%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in SEQ ID NO:2, or a portion thereof. The polynucleotide of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions as defined herein, to one of the nucleotide sequences shown in SEQ ID NO:2, or a portion thereof.

Moreover, the polynucleotide of the invention can comprise only a portion of the coding region of one of the sequences in SEQ ID NO:2, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an acetyl-CoA carboxylase. The nucleotide sequences determined from the cloning of the acetyl-CoA carboxylase gene from *P. rhodozyma* allows for the generation of probes and primers designed for use in identifying and/or cloning acetyl-CoA carboxylase homologues in other cell types and organisms. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 15 preferably about 20 or 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth, e.g., in SEQ ID NO: No:2, an anti-sense sequence of one of the sequences, e.g., set forth in SEQ ID NO:2, or naturally occurring mutants thereof. Primers based on a nucleotide of invention can be used in PCR reactions to clone acetyl-CoA carboxylase homologues. Probes based on the acetyl-CoA carboxylase nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. The probe can further comprise a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a genomic marker test kit for identifying cells which express an acetyl-CoA carboxylase, such as by measuring a level of

an acetyl-CoA carboxylase-encoding nucleic acid molecule in a sample of cells, e.g., detecting acetyl-CoA carboxylase mRNA levels or determining whether a genomic acetyl-CoA carboxylase gene has been mutated or deleted.

The polynucleotide of the invention encodes a polypeptide or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of SEQ ID NO:3 such that the protein or portion thereof maintains an acetyl-CoA carboxylase activity, in particular an acetyl-CoA carboxylase activity as described in the examples in microorganisms or plants. As used herein, the language "sufficiently homologous" refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of the polypeptide of the present invention amino acid residues to an amino acid sequence of SEQ ID NO:3 such that the protein or portion thereof has an acetyl-CoA carboxylase activity. Examples of an acetyl-CoA carboxylase activity are also described herein.

The protein is at least about 60-65%, preferably at least about 66-70%, and more preferably at least about 70- 80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of SEQ ID NO:3.

Portions of proteins encoded by the acetyl-CoA carboxylase polynucleotide of the invention are preferably biologically active portions of one of the acetyl-CoA carboxylase.

As mentioned herein, the term "biologically active portion of acetyl-CoA carboxylase" is intended to include a portion, e.g., a domain/motif, that has acetyl-CoA carboxylase activity or has an immunological activity such that it is binds to an antibody binding specifically to acetyl-CoA carboxylase. To determine whether an acetyl-CoA carboxylase or a biologically active portion thereof can participate in the metabolism an assay of enzymatic activity may be performed. Such assay methods are well known to those skilled in the art, as detailed in the Examples. Additional nucleic acid fragments encoding biologically active portions of an acetyl-CoA carboxylase can be prepared by isolating a portion of one of the sequences in SEQ ID NO:2, expressing the encoded portion of the acetyl-CoA carboxylase or peptide (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the acetyl-CoA carboxylase or peptide.

At first, a partial gene fragment was cloned containing a portion of the ACC gene by using the degenerate PCR method. Said degenerate PCR is a method to clone a gene of interest which has high homology of amino acid sequence to the known enzyme from other species

which has the same or similar function. Degenerate primer, which is used as a primer in degenerate PCR, was designed by a reverse translation of the amino acid sequence to corresponding nucleotides ("degenerated"). In such a degenerate primer, a mixed primer which consists any of A, C, G or T, or a primer containing inosine at an ambiguity code is generally used. In this invention, such mixed primers were used for degenerate primers to clone above gene.

An entire gene containing its coding region with its intron as well as its regulation region such as a promoter or a terminator can be cloned from a chromosome by screening of a genomic library which is constructed in phage vector or plasmid vector in appropriate host, by using a partial DNA fragment obtained by degenerate PCR as described above as a probe after it was labeled. Generally, *E. coli* as a host strain and *E. coli* vector, a phage vector such as λ phage vector, or a plasmid vector such as pUC vector is often used in the construction of a library and a following genetic manipulation such as a sequencing, a restriction digestion, a ligation and the like. In this invention, an *EcoR*I genomic library of *P. rhodozyma* was constructed in the derivatives of λ vector, λZAPII. An insert size, what length of insert must be cloned, was determined by the Southern blot hybridization for the gene before construction of a library. In this invention, a DNA used for a probe was labeled with digoxigenin (DIG), a steroid hapten instead of conventional ^{32}P label, following the protocol which was prepared by the supplier (Boehringer-Mannheim, Mannheim, Germany). A genomic library constructed from the chromosome of *P. rhodozyma* was screened by using a DIG-labeled DNA fragment which had a portion of a gene of interest as a probe. Hybridized plaques were picked up and used for further study. When λZAPII (insert size was below 9kb) was used in the construction of the genomic library, in vivo excision protocol was conveniently used for the succeeding step of the cloning into the plasmid vector by using a derivative of single stranded M13 phage, Ex assist phage (Stratagene, La Jolla, USA). A plasmid DNA thus obtained was examined for sequencing.

In this invention, we used the automated fluorescent DNA sequencer, ALFred system (Pharmacia, Uppsala, Sweden) using an autocycle sequencing protocol in which the Taq DNA polymerase is employed in most cases of sequencing.

After the determination of the genomic sequence, a sequence of a coding region was used for a cloning of cDNA of corresponding gene. The PCR method was also exploited to clone cDNA fragment. The PCR primers whose sequences were identical to the sequence at the 5'- and 3'- end of the open reading frame (ORF) were synthesized with an addition of an appropriate restriction site, and PCR was performed by using those PCR primers. In

this invention, a cDNA pool was used as a template in this PCR cloning of cDNA. The said cDNA pool consists of various cDNA species which were synthesized *in vitro* by the viral reverse transcriptase and Taq polymerase (CapFinder Kit manufactured by Clontech, Palo Alto, U.S.A.) by using the mRNA obtained from *P. rhodozyma* as a template. cDNA of interest thus obtained was confirmed in its sequence. Furthermore, cDNA thus obtained was used for a confirmation of its enzyme activity after the cloning of the cDNA fragment into an expression vector which functions in *E. coli* under the strong promoter activity such as the *lac* or T7 expression system.

In another embodiment, the present invention relates to a method for making a recombinant vector comprising inserting a polynucleotide of the invention into a vector.

Further, the present invention relates to a recombinant vector containing the polynucleotide of the invention or produced by said method of the invention.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting a polynucleotide to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA or PNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The present invention also relates to cosmids, viruses, bacteriophages and other vectors used conventionally in genetic engineering that contain a nucleic acid molecule according to the invention. Methods which are well known to those skilled in the art can be used to construct various plasmids and vectors. Alternatively, the nucleic acid molecules and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

The present invention further relates to a vector in which the polynucleotide of the present invention is operatively linked to expression control sequences allowing expression in prokaryotic or eukaryotic host cells. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoter, ribosomal binding site, and terminators. In eukaryotes, generally control sequences include promoters, terminators and, in some instances, enhancers, transactivators; or transcription factors.

The term "control sequence" is intended to include, at a minimum, components the presence of which are necessary for expression, and may also include additional advantageous components.

The term "operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. In case the control sequence is a promoter, it is obvious for a skilled person that double-stranded nucleic acid is used.

Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells or under certain conditions. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by polynucleotides as described herein.

The recombinant expression vectors of the invention can be designed for expression of acetyl-CoA carboxylase in prokaryotic or eukaryotic cells. For example, genes encoding the polynucleotide of the invention can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast and other fungal cells, algae, ciliates of the types: *Holotrichia, Peritrichia, Spirotrichia, Suctoria, Tetrahymena, Paramecium, Colpidium, Glaucoma, Platyophrya, Potomacus, Pseudocohnilembus, Euplotes, Engelmaniella,* and *Stylonychia*, especially *Stylonychia lemnae* with vectors following, a transformation method as described in WO9801572 and multicellular plant cells. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc.), pMAL (New England Biolabs, Beverly, USA) and pRIT5 (Pharmacia, Piscataway, USA) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the polypeptide encoded by the polynucleotide of the present invention is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin, e.g. recombinant acetyl-CoA carboxylase unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc and pET 11d. Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 *gn10*-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 *gnl*). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 *gn1* gene under the transcriptional control of the *lac*UV 5 promoter.

One strategy to maximize recombinant protein expression is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially

utilized in the bacterium chosen for expression, such as *E. coli*. Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

Further, the acetyl-CoA carboxylase vector can be a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1, pMFa, pJRY88, and pYES2 (Invitrogen, San Diego, USA). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, are known to the skilled artisan.

Alternatively, the polynucleotide of the invention can be introduced in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series and the pVL series.

Alternatively, the polynucleotide of the invention is introduced in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 and pMT2PC. When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

The recombinant mammalian expression vector can be capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue- specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver- specific), lymphoid-specific promoters, in particular promoters of T cell receptors and immunoglobulins, neuron-specific promoters (e.g., the neurofilament promoter), pancreas-specific promoters, and mammary gland-specific promoters (e.g., milk whey promoter; US 4,873, 316 and EP 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters and the fetoprotein promoter.

Thus expressed ACC gene can be verified for its activity, e.g., by an enzyme assay method. Some experimental protocols are described in the literature. The following is the one of the methods which is used for the determination of acetyl-CoA carboxylase activity: Assays are performed by measuring the loss in acetyl-CoA and/or the production of malonyl-CoA at 5 min intervals for 20 min, using reverse phase HPLC. The rate of conversion of acetyl-CoA to malonyl-CoA is found to be linear for 20 min, and velocities are calculated by linear regression analysis of the malonyl-CoA concentration with respect to time. The

reaction mixture contained 50 mM Tris, pH 7.5, 6 µM acetyl-CoA, 2 mM ATP, 7 mM $KHCO_3$, 8 mM $MgCl_2$, 1 mM dithiothreitol, and 1 mg/ml bovine serum albumin. Enzyme is preincubated (30 min, 25°C) with bovine serum albumin (2 mg/ml) and potassium citrate (10 mM). Reactions are initiated by transferring 50 µl of preincubated enzyme to the reaction mixture (final volume 200 µl) and incubated for 5-20 min at 25°C. Reactions are terminated by addition of 50 µl 10% perchloric acid. Following termination of the reaction, the samples are centrifuged (3 min, 10,000 × g) and analyzed by HPLC. A mobile phase of 10 mM KH_2PO_4, pH 6.7 (solvent A), and MeOH (solvent B) is used. The flow rate is 1.0 ml/min, and the gradient is as follows: hold at 100% solvent A for 1 min followed by a linear gradient to 30% solvent B over the next 5 min, then hold at 30% solvent B for 5 min. Using this method the retention times were 7.5 and 9.0 min for malonyl-CoA and acetyl-CoA, respectively. When an expression vector for *S. cerevisiae* is used, a complementation analysis can be conveniently exploited by using conditional acetyl-CoA carboxylase null mutant strain derived from *S. cerevisiae* as a host strain for its confirmation of activity.

Succeeding to the confirmation of the enzyme activity, an expressed protein would be purified and used for raising the antibody against the purified enzyme. Antibody thus prepared would be used for a characterization of the expression of the corresponding enzyme in a strain improvement study, an optimization study of the culture condition, and the like.

In a further embodiment, the present invention relates to an antibody that binds specifically to the polypeptide of the present invention or parts, i.e. specific fragments or epitopes of such a protein.

The antibodies of the invention can be used to identify and isolate other acetyl-CoA carboxylase and genes. These antibodies can be monoclonal antibodies, polyclonal antibodies or synthetic antibodies as well as fragments of antibodies, such as Fab, Fv or scFv fragments etc. Monoclonal antibodies can be prepared, for example, by the techniques as originally described by Kohler and Milstein, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals.

Furthermore, antibodies or fragments thereof to the aforementioned peptides can be obtained by using methods known to the skilled person. These antibodies can be used, for example, for the immunoprecipitation and immunolocalization of proteins according to the invention as well as for the monitoring of the synthesis of such proteins, for example, in recombinant organisms, and for the identification of compounds interacting with the

protein according to the invention. For example, surface plasmon resonance as employed in the BIAcore system can be used to increase the efficiency of phage antibodies selections, yielding a high increment of affinity from a single library of phage antibodies which bind to an epitope of the protein of the invention. In many cases, the binding phenomenon of antibodies to antigens is equivalent to other ligand/anti-ligand binding.

In this invention, the gene fragment for acetyl-CoA carboxylase was cloned from *P. rhodozyma* with a purpose to decrease its expression level in *P. rhodozyma* by genetic method using the cloned gene fragment.

To decrease a gene expression with genetic methods, some strategies can be employed, one of which is a gene-disruption method. In this method, a partial fragment of the objective gene to be disrupted is ligated to a drug resistant cassette on the integration vector which can not replicate in the host organism. A drug resistance gene which encodes the enzyme that enables the host to survive in the presence of a toxic antibiotic is often used for the selectable marker. G418 resistance gene harbored in pGB-Ph9 (Wery *et al.* (Gene, 184, 89-97, 1997)) is an example of a drug resistance gene which functions in *P. rhodozyma.* Nutrition complementation marker can be also used in the host which has an appropriate auxotrophy marker. *P. rhodozyma* ATCC24221 strain that requires cytidine for its growth is one example of the auxotroph. By using CTP synthetase as donor DNA for ATCC24221, a host vector system using a nutrition complementation can be established.

After the transformation of the host organisms and recombination between the objective gene fragment on the vector and its corresponding gene fragment on the chromosome of the host organisms, the integration vector is integrated onto the host chromosome by single cross recombination. As a result of this recombination, the drug resistant cassette would be inserted in the objective gene whose translated product is only synthesized in its truncated form which does not have its enzymatic function. In a similar manner, two parts of the objective gene were also used for gene disruption study in which the drug resistant gene can be inserted between such two partial fragments of the objective genes on the integration vector. In the case of this type of vector, double recombination event between the gene fragments harbored on the integration vector and the corresponding gene fragments on the chromosome of the host are expected. Although frequency of this double crossing-over recombination is lower than single cross recombination, null phenotype of the objective gene by the double cross recombination is more stable than by the single cross recombination.

On the other hand, this strategy has difficulty in the case of the gene whose function is essential and disruption is lethal for the host organism such as acetyl-CoA carboxylase gene. The function of acetyl-CoA carboxylase is indispensable for the host survival other than the biosynthesis of fatty acid. From such a viewpoint, it seemed to be difficult to construct the acetyl-CoA carboxylase disruptant from *P. rhodozyma* by this gene disruption method.

In such a case, other strategies can be applied to decrease (not to disrupt) a gene expression, one of which is a conventional mutagenesis to screen the mutant whose expression for acetyl-CoA carboxylase is decreased. In this method, an appropriate recombinant in which an appropriate reporter gene is fused to the promoter region of acetyl-CoA carboxylase gene from the host organism is mutated and mutants which show a weaker activity of reporter gene product can be screened. In such mutants, it is expected that their expression of acetyl-CoA carboxylase activity decreased by the mutation lying in the promoter region of reporter gene or *trans*-acting region which might affect the expression of acetyl-CoA carboxylase gene other than the mutation lying in the promoter gene itself. In the case of mutation occurring at the promoter region of the reporter fusion, such mutation can be isolated by the sequence of the corresponding region. Thus isolated mutation can be introduced in a variety of carotenoids, especially astaxanthin producing mutants derived from *P. rhodozyma* by a recombination between the original promoter for acetyl-CoA carboxylase gene on the chromosome and the mutated promoter fragment. To exclude mutations occurring at a *trans*-acting region, a mutation can also be induced by an *in vitro* mutagenesis of a *cis* element in the promoter region. In this approach, a gene cassette, containing a reporter gene which is fused to a promoter region derived from a gene of interest at its 5'-end and a terminator region from a gene of interest at its 3'-end, is mutagenized and then introduced into *P. rhodozyma*. By detecting the difference of the activity of the reporter gene, an effective mutation can be screened. Such a mutation can be introduced in the sequence of the native promoter region on the chromosome by the same method as the case of an *in vivo* mutation approach. But, these methods have some drawbacks to have some time-consuming process.

Another strategy to decrease a gene expression is an antisense method. This method is frequently applied to decrease the gene expression even when teleomorphic organisms such as *P. rhodozyma* are used as host organisms, to which the mutation and gene disruption method is usually difficult to be applied. The anti-sense method is a method to decrease an expression of gene of interest by introducing an artificial gene fragment, whose sequence is complementary to cDNA fragment of the gene of interest. Such an anti-sense

gene fragment would form a complex with a mature mRNA fragment of the objective gene *in vivo* and inhibit an efficient translation from mRNA, as a consequence.

An "antisense" nucleic acid molecule comprises a nucleotide sequence which is complementary to a "sense" nucleic acid molecule encoding a protein, e. g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to a mRNA sequence. Accordingly, an antisense nucleic acid molecule can hydrogen bond to a sense nucleic acid molecule. The antisense nucleic acid molecule can be complementary to an entire acetyl-CoA carboxylase-coding strand, or to only a portion thereof. Accordingly, an antisense nucleic acid molecule can be antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an acetyl-CoA carboxylase. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. Further, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding acetyl-CoA carboxylase. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into a polypeptide (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding acetyl-CoA carboxylase disclosed herein, antisense nucleic acid molecules of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of acetyl-CoA carboxylase mRNA, but can also be an oligonucleotide which is antisense to only a portion of the coding or noncoding region of acetyl-CoA carboxylase mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of acetyl-CoA carboxylase mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid molecule of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid molecule (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the anti-sense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-

D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a polynucleotide has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted polynucleotide will be of an antisense orientation to a target polynucleotide of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a cell or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an acetyl-CoA carboxylase to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The anti-sense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic including plant promoters are preferred.

The antisense nucleic acid molecule of the invention may, e.g., be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide or a chimeric RNA-DNA analogue.

Further the antisense nucleic acid molecule of the invention can be a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a

single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes) can be used to catalytically cleave acetyl-CoA carboxylase mRNA transcripts to thereby inhibit translation of mRNA. A ribozyme having specificity for an acetyl-CoA carboxylase-encoding nucleic acid molecule can be designed based upon the nucleotide sequence of an acetyl-CoA carboxylase cDNA disclosed herein or on the basis of a heterologous sequence to be isolated according to methods taught in this invention. For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an encoding mRNA (see, e.g., US 4,987,071 and US 5,116,742). Alternatively, acetyl-CoA carboxylase mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules.

The application of the antisense method to construct a carotenoid overproducing strain from *P. rhodozyma* is disclosed in EP 1,158,051.

In one embodiment the present invention relates to a method of making a recombinant host cell comprising introducing the vector or the polynucleotide of the present invention into a host cell.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection", conjugation and transduction are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-mediated transfer, or electroporation. Suitable methods for transforming or transfecting host cells including plant cells are known to the skilled artisan.

For stable transfection of mammalian cells, only a small fraction of cells may integrate the foreign DNA into their genome, depending upon the expression vector and transfection technique used. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding the polypeptide of the present invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug

selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of the polynucleotide of the present invention into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the acetyl-CoA carboxylase gene. Preferably, this acetyl-CoA carboxylase gene is a *P. rhodozyma* acetyl-CoA carboxylase gene, but it can be a homologue from a related or different source. Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous acetyl-CoA carboxylase gene is mutated or otherwise altered but still encodes a functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous acetyl-CoA carboxylase). To create a point mutation via homologous recombination also DNA-RNA hybrids can be used known as chimeraplasty known from Cole-Strauss *et al.*, Nucl. Aci. Res., 27, 5, 1323-1330, 1999 and Kmiec, Gene therapy., American Scientist. 87, 3, 240-247. 1999.

The vector is introduced into a cell and cells in which the introduced polynucleotide gene has homologously recombined with the endogenous acetyl-CoA carboxylase gene are selected, using art-known techniques.

Further host cells can be produced which contain selection systems which allow for regulated expression of the introduced gene. For example, inclusion of the polynucleotide of the invention on a vector placing it under control of the lac operon permits expression of the polynucleotide only in the presence of IPTG. Such regulatory systems are well known in the art.

Preferably, the introduced nucleic acid molecule is foreign to the host cell.

By "foreign" it is meant that the nucleic acid molecule is either heterologous with, respect to the host cell, this means derived from a cell or organism with a different genomic background, or is homologous with respect to the host cell but located in a different genomic environment than the naturally occurring counterpart of said nucleic acid molecule. This means that, if the nucleic acid molecule is homologous with respect to the host cell, it is not located in its natural location in the genome of said host cell, in particular it is surrounded by different genes. In this case the nucleic acid molecule may be either under the control of its own promoter or under the control of a heterologous promoter. The vector or nucleic acid molecule according to the invention which is present in the host cell may either be integrated into the genome of the host cell or it may be maintained in some form

extrachromosomally. In this respect, it is also to be understood that the nucleic acid molecule of the invention can be used to restore or create a mutant gene via homologous recombination.

Accordingly, in another embodiment the present invention relates to a host cell genetically engineered with the polynucleotide of the invention or the vector of the invention.

The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

For example, a polynucleotide of the present invention can be introduced in bacterial cells as well as insect cells, fungal cells or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells), algae, ciliates, plant cells, fungi or other microorganims like *E. coli*. Other suitable host cells are known to those skilled in the art. Preferred are *E. coli*, baculovirus, Agrobacterium or fungal cells are, for example, those of the genus *Saccharomyces*, e.g. those of the species *S. cerevisiae* or *P. rhodozyma* (*Xanthophylomyces dendrorhous*).

In addition, in one embodiment, the present invention relates to a method for the production of fungal transformants comprising the introduction of the polynucleotide or the vector of the present invention into the genome of said fungal cell.

For the expression of the nucleic acid molecules according to the invention in sense or antisense orientation in plant cells, the molecules are placed under the control of regulatory elements which ensure the expression in fungal cells. These regulatory elements may be heterologous or homologous with respect to the nucleic acid molecule to be expressed as well with respect to the fungal species to be transformed.

In general, such regulatory elements comprise a promoter active in fungal cells. To obtain constitutive expression in fungal cells, preferably constitutive promoters are used, e.g., the glyceraldehyde-3-dehydrogenase promoter derived from *P. rhodozyma* (WO 97/23,633). Inducible promoters may be used in order to be able to exactly control expression. An example for inducible promoters is the promoter of genes encoding heat shock proteins. Also an amylase gene promoter which is a candidate for such inducible promoters has been described (EP 1,035,206). The regulatory elements may further comprise transcriptional and/or translational enhancers functional in fungal cells. Furthermore, the regula-

tory elements may include transcription termination signals, such as a poly-A signal, which lead to the addition of a poly A tail to the transcript which may improve its stability.

Methods for the introduction of foreign DNA into fungal cells are also well known in the art. These include, for example, transformation with the LiCl method, the fusion of protoplasts, electroporation, biolistic methods like particle bombardment other methods known in the art. Methods for the transformation using biolistic methods are well known to the person skilled in the art.

The term "transformation" as used herein, refers to the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for the transfer. The polynucleotide may be transiently or stably introduced into the host cell and may be maintained non- integrated, for example, as a plasmid or as chimeric links, or alternatively, may be integrated into the host genome.

In general, the fungi which can be modified according to the invention and which either show overexpression of a protein according to the invention or a reduction of the synthesis of such a protein can be derived from any desired fungal species.

Further, in one embodiment, the present invention relates to a fungal cell comprising the polynucleotide the vector or obtainable by the method of the present invention.

Thus, the present invention relates also to transgenic fungal cells which contain (preferably stably integrated into the genome) a polynucleotide according to the invention linked to regulatory elements which allow expression of the polynucleotide in fungal cells and wherein the polynucleotide is foreign to the transformed fungal cell. For the meaning of foreign; see supra.

Thus, the present invention also relates to transformed fungal cells according to the invention.

Accordingly, due to the altered expression of acetyl-CoA carboxylase, cells metabolic pathways are modulated in yield production, and/or efficiency of production.

The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example fatty acids, carotenoids, (poly)saccharides, lipids, vitamins, isoprenoids, wax esters, and/or polymers like polyhydroxyalkanoates and/or its metabolism products or further desired fine chemical as mentioned herein) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter).

The term "efficiency" of production includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a said altered yield, in particular, into carotenoids, (poly)saccharides, lipids, vitamins, isoprenoids etc.).

The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e. acetyl CoA, fatty acids, vitamins, carotenoids, isoprenoids, lipids etc. and/or further compounds as defined above and which biosynthesis is based on said products). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased.

The terms "biosynthesis" (which is used synonymously for "synthesis" of "biological production" in cells, tissues plants, etc.) or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process.

The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of acetyl CoA, a fatty acid, hexose, isoprenoid, vitamin, carotenoid, lipid etc.) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

Such a genetically engineered *P. rhodozyma* would be cultivated in an appropriate medium and evaluated in its productivity of carotenoids, especially astaxanthin. A hyper producer of astaxanthin thus selected would be confirmed in view of the relationship between its productivity and the level of gene or protein expression which is introduced by such a genetic engineering method.

The present invention is further illustrated with Examples described below.

The following materials and methods employed in the Examples are described below:

Strains

P. rhodozyma ATCC96594 (re-deposited under the accession No. ATCC 74438 on April 8, 1998 pursuant to the Budapest Treaty)

E. coli DH5α: F^-, ϕ80d, *lac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *hsd* (r_K^-, m_K^+), *rec*A1, *end*A1, *deo*R, *thi*-1, *sup*E44, *gyr*A96, *rel*A1 (Toyobo, Osaka, Japan)

E. coli XL1-Blue MRF': Δ(*mcr*A)183, Δ(*mcr*CB-*hsd*SMR-*mrr*)173, *end*A1, *sup*E44, *thi*-1, *rec*A1, *gyr*A96, *rel*A1, *lac* [F' *pro*AB, *lac*IqZΔM15, Tn10 (tet^r)] (Stratagene, La Jolla, USA)

E. coli SOLR: *e*14-(*mcr*A), Δ(*mcr*CB-*hsd*SMR-*mrr*)171, *sbc*C, *rec*B, *rec*J, *umu*C :: Tn5(kan^r), *uvr*C, *lac*, *gyr*A96, *rel*A1, *thi*-1, *end*A1, ΔR, [F' *pro*AB, *lac*IqZ ΔM15] Su-(nonsuppressing) (Stratagene)

E. coli TOP10: F-, *mcr*A, Δ*mrr-hsd*RMS-*mcr*BC), *ϕ80*, Δ*lac*Z M15, Δ*lac*X74, *rec*A1, *deo*R, *ara*D139, (*ara-leu*)7697, *gal*U, *gal*K, rpsL (Str^r), *end*A1, *nup*G (Invitrogen, Carlsbad, USA)

Vectors

λZAPII (Stratagene)

pBluescriptII KS- (Stratagene)

pMOSBlue T-vector (Amersham, Buckinghamshire, U.K.)

pCR2.1-TOPO (Invitrogen)

Media

P. rhodozyma strain was maintained routinely in YPD medium (DIFCO, Detroit, U.S.A.). *E. coli* strain was maintained in LB medium (10 g Bacto-trypton, 5 g yeast extract (DIFCO) and 5 g NaCl per liter). NZY medium (5 g NaCl, 2 g $MgSO_4$-$7H_2O$, 5 g yeast extract (DIFCO), 10 g NZ amine type A (WAKO, Osaka, Japan) per liter) is used for λ phage propagation in a soft agar (0.7 % agar (WAKO)). When an agar medium was prepared, 1.5 % of agar (WAKO) was supplemented.

Methods

Restriction enzymes and T4 DNA ligase were purchased from Takara Shuzo (Ohtsu, Japan).

Isolation of a chromosomal DNA from *P. rhodozyma* was performed by using QIAGEN Genomic Kit (QIAGEN, Hilden, Germany) following the protocol supplied by the manufacturer. Mini-prep of plasmid DNA from transformed *E. coli* was performed with the Automatic DNA isolation system (PI-50, Kurabo, Co. Ltd., Osaka, Japan). Midi-prep of plasmid DNA from an *E. coli* transformant was performed by using QIAGEN column (QIAGEN). Isolation of λ DNA was performed by Wizard lambda preps DNA purification system (Promega, Madison, U.S.A.) following the protocol prepared by the manufacturer. A DNA fragment was isolated and purified from agarose by using QIAquick or QIAEX II (QIAGEN). Manipulation of λ phage derivatives was followed by the protocol prepared by the manufacturer (Stratagene).

Isolation of total RNA from *P. rhodozyma* was performed with the phenol method by using Isogen (Nippon Gene, Toyama, Japan). mRNA was purified from total RNA thus obtained by using mRNA separation kit (Clontech). cDNA was synthesized by using CapFinder cDNA construction kit (Clontech).

In vitro packaging was performed by using Gigapack III gold packaging extract (Stratagene).

The polymerase chain reaction (PCR) was performed with the thermal cycler from Perkin Elmer model 2400. Each PCR condition is described in examples. PCR primers were purchased from a commercial supplier. Fluorescent DNA primers for DNA sequencing were purchased from Pharmacia. DNA sequencing was performed with the automated fluorescent DNA sequencer (ALFred, Pharmacia).

Competent cells of DH5α were purchased from Toyobo (Japan).

Example 1: Isolation of mRNA from *P. rhodozyma* and construction of cDNA library

To construct cDNA library of *P. rhodozyma*, total RNA was isolated by phenol extraction method right after the cell disruption and the mRNA from *P. rhodozyma* ATCC96594 strain was purified by using mRNA separation kit (Clontech).

At first, Cells of ATCC96594 strain from 10 ml of two-day-culture in YPD medium were harvested by centrifugation (1500 x g for 10 min.) and washed once with extraction buffer (10 mM Na-citrate / HCl (pH 6.2) containing 0.7 M KCl). After suspending in 2.5 ml of extraction buffer, the cells were disrupted by French press homogenizer (Ohtake Works Corp., Tokyo, Japan) at 1500 kgf/cm2 and immediately mixed with two times of volume of isogen (Nippon gene) according to the method specified by the manufacturer. In this step, 400 μg of total RNA was recovered.

Then, this total RNA was purified by using mRNA separation kit (Clontech) according to the method specified by the manufacturer. Finally, 16 μg of mRNA from *P. rhodozyma* ATCC96594 strain was obtained.

To construct cDNA library, CapFinder PCR cDNA construction kit (Clontech) was used according to the method specified by the manufacturer. One μg of purified mRNA was applied for a first strand synthesis followed by PCR amplification. After this amplification by PCR, 1 mg of cDNA pool was obtained.

Example 2: Cloning of a partial *ACC* (acetyl-CoA carboxylase) gene from *P. rhodozyma*

To clone a partial *ACC* gene from *P. rhodozyma*, a degenerate PCR method was exploited. Species and accession number to database whose sequence for acetyl-CoA carboxylase were used for multiple alignment analysis are as follows.

Arabidopsis thaliana	D34630 (DDBJ)
Emericella nidulans	Y15996 (EMBL)
Gallus gallus	P11029 (Swiss-Prot)
Glycine max	L48995 (GenBank)
Homo sapiens	S41121 (PIR)
Medicago sativa	L25042 (GenBank)
Ovis aries	Q28559 (Swiss-Prot)
Rattus norvegicus	P11497 (Swiss-Prot)
Saccharomyces cerevisiae	Q00955 (Swiss-Prot)
Schizosaccharomyces pombe	P78820 (Swiss-Prot)
Ustilago maydis	S49991 (PIR)

Two mixed primers whose nucleotide sequences were designed and synthesized based on the common sequence of known acetyl-CoA carboxylase genes from other species: acc9 (sense primer) (SEQ ID NO:4) and acc13 (antisense primer) (SEQ ID NO:5) (in the sequences "n" means nucleotides a, c, g or t, "h" means nucleotides a, c or t, "m" means nucleotides a or c, "k" means nucleotides g or t, and "y" means nucleotides c or t). After the PCR reaction of 25 cycles of 95°C for 30 seconds, 45°C for 30 seconds and 72°C for 15 seconds by using ExTaq (Takara Shuzo) as a DNA polymerase and cDNA pool obtained in Example 1 as a template, reaction mixture was applied to agarose gel electrophoresis. One PCR band that had a desired length (0.8 kb) was recovered from the agarose gel and purified by QIAquick (QIAGEN) according to the method by the manufacturer and then ligated to pMOSBlue-T-vector (Amersham). After transformation of competent *E. coli* DH5α, 6 white colonies were selected and plasmids were isolated with Automatic DNA isolation system. As a result of sequencing, it was found that 3 clones had a sequence whose deduced amino acid sequence was similar to known acetyl-CoA carboxylase genes. These isolated cDNA clones were designated as pACC1014 and used for further screening study.

Example 3: Isolation of genomic DNA from *P. rhodozyma*

To isolate a genomic DNA from *P. rhodozyma*, QIAGEN genomic kit was used according to the method specified by the manufacturer.

At first, cells of *P. rhodozyma* ATCC96594 strain from 100 ml of overnight culture in YPD medium were harvested by centrifugation (1500 x g for 10 min.) and washed once with TE buffer (10 mM Tris / HCl (pH 8.0) containing 1 mM EDTA). After suspending in 8 ml of Y1 buffer of the QIAGEN genomic kit, lyticase (SIGMA, St. Louis, U.S.A.) was added at the concentration of 2 mg/ml to disrupt cells by enzymatic degradation and the reaction mixture was incubated for 90 min at 30°C and then proceeded to the next extraction step. Finally, 20 µg of genomic DNA was obtained.

Example 4: Southern blot hybridization by using pACC1014 as a probe

Southern blot hybridization was performed to clone a genomic fragment which contains *ACC* gene from *P. rhodozyma*. Two µg of genomic DNA was digested by *Eco*RI and subjected to agarose gel electrophoresis followed by acidic and alkaline treatment. The denatured DNA was transferred to nylon membrane (Hybond N+, Amersham) by using transblot (Joto Rika, Tokyo, Japan) for an hour. The DNA which was transferred to nylon membrane was fixed by a heat treatment (80°C, 90 min). A probe was prepared by labeling a template DNA (*Eco*RI and *Sal*I -digested pACC1014) with DIG multipriming method (Boehringer Mannheim). Hybridization was performed with the method specified by the manufacturer. As a result, a hybridized band was visualized in the range from 2.0 to 2.3 kilobases (kb).

Example 5: Cloning of a genomic fragment containing the *ACC* gene

4 µg of the genomic DNA were digested by *Eco*RI and subjected to agarose gel electrophoresis. Then, DNAs with a length within the range from 1.5 to 2.7 kb was recovered by QIAEX II gel extraction kit (QIAGEN) according to the method specified by the manufacturer. The purified DNA was ligated to 0.5 µg of *Eco*RI-digested and CIAP (calf intestine alkaline phosphatase)-treated λZAP II (Stratagene) at 16°C overnight, and packaged by Gigapack III gold packaging extract (Stratagene). The packaged extract was infected to *E. coli* MRF' strain and over-laid with NZY medium poured onto LB agar medium. About 5000 plaques were screened by using *Eco*RI and *Sal*I-digested pACC1014 as a probe. Five plaques were hybridized to the labeled probe.

The *in vivo* excision protocol was applied to these λZAP II derivatives containing putative *ACC* gene from *P. rhodozyma* by following the instruction manual (Stratagene) to clone the insert fragment into *E. coli* cloning vector, pBluescript SK. Each clone recovered from five positive plaques was subjected for sequencing analysis and it was found that the three of them had the identical sequence to the insert fragment of pACC1014. One of the clone

was named as pACC1224 and used for further study. As a result of whole sequencing of the entire region of insert fragment in pACC1224, it was suggested that this clone contained neither its 5'- nor 3'-end of the *ACC* gene.

Example 6: Cloning of the flanking region of the insert fragment in pACC1224 from the genome of *P. rhodozyma* by genome walking method

Two PCR primers were synthesized based on the internal sequence of pACC1224 and used for the genome walking method: acc17 (SEQ ID NO:6) and acc18 (SEQ ID NO:7). The protocol of the instruction manual provided from the supplier (Clontech) was followed for the genome walking method. In the PCR reaction using acc17 primer, a 2.8 kb PCR band emerged from the genomic *Stu*I library. In the case of acc18 primer, a 2.2 kb PCR band was produced in the genomic *Pvu*II library. These PCR bands were cloned into pCR2.1-TOPO (Invitrogen) and it was revealed that 2.8 kb PCR band contained a 5' fragment of *ACC* gene and 2.2 kb PCR band contained 3' fragment of *ACC* gene, respectively. The clones containing 2.8 kb and 2.2 kb PCR fragment were named as pACCStu107 and pACCPvd107, respectively and used for further study.

Example 7: Southern blot hybridization by using pACCStu107 and pACCPvd107 as probes

Southern blot hybridization was performed to clone a genomic fragment which covered the *ACC* gene from *P. rhodozyma*. 2 µg of genomic DNA was digested by *Eco*RI and subjected to agarose gel electrophoresis followed by acidic and alkaline treatment. The denatured DNA was transferred to nylon membrane (Hybond N+, Amersham) by using transblot (Joto Rika, Tokyo, Japan) for an hour. The DNA which was transferred to nylon membrane was fixed by a heat treatment (80°C, 90 min). A probe was prepared by labeling a template DNA (*Eco*RI -digested pACCStu107 and pACCPvd107) with the DIG multi-priming method (Boehringer Mannheim). Hybridization was performed with the method specified by the manufacturer. As a result, several hybridized bands whose size was close to 2.0 kb, 0.9 kb and 0.6 kb were visualized when the insert fragment in pACCStu107 was used as a probe. In the case that the insert fragment in pACCPvd107 was used as a probe, a hybridized band was visualized in the range from 6.0 kb to 6.5 kb.

Example 8: Cloning of the genomic clone covering the *ACC* gene

In a similar manner to Example 5, the genomic fragment containing the insert fragment in pACCStu107 and pACCPvd107 was cloned by plaque hybridization. 4 µg of the genomic

DNA was digested by *Eco*RI and subjected to agarose gel electrophoresis. Then, DNAs with a length within the following range were recovered by QIAEX II gel extraction kit (QIAGEN) according to the method specified by the manufacturer: (1) from 2.7 to 5.0 kb; (2) from 1.4 to 2.7 kb; and (3) from 0.5 to 1.4 kb.

Each purified DNA was ligated to 0.5 µg of *Eco*RI-digested and CIAP (calf intestine alkaline phosphatase)-treated λZAP II (Stratagene) at 16 °C overnight, and packaged by Gigapack III gold packaging extract (Stratagene). The packaged extract was infected to *E. coli* MRF' strain and over-laid with NZY medium poured onto LB agar medium. About 5000 plaques were screened by using *Eco*RI -digested pACCStu107 and pACCPvd107 as probes.

The following candidates were isolated after plaque hybridization study.

1) 3 plaques from the 2.7 to 6.0 kb library by using the insert of pACCPvd107 as a probe.

2) 3 plaques from the 1.4 to 2.7 kb library by using the insert of pACCStu107 as a probe.

3) 21 plaques from the 0.5 to 1.4 kb library by using the insert of pACCStu107 as a probe.

The *in vivo* excision protocol was applied to these λZAP II derivatives containing putative *ACC* gene from *P. rhodozyma* by following the instruction manual (Stratagene) to clone the insert fragment into *E. coli* cloning vector, pBluescript SK. Each clone recovered from the positive plaques was subjected for sequencing analysis. At least each clone had the putative *ACC* gene from BLAST X analysis (http://www.blast.genome.ad.jp/). The following clones were selected and used for further analysis:

pACC119-18 having a 6 kb insert and covering the 3' end of the *ACC* gene;

pACC119-17-0.6 having a 0.6 kb insert flanking the 5' end of the pACC1224 insert fragment;

pACC119-17-2 having a 2 kb insert flanking the 5' end of the pACC119-17-0.6 insert fragment; and

pACC127-17-0.9 having a 0.9kb insert flanking the 5' end of the pACC119-17-2 insert fragment.

As a result of whole sequencing of the entire region of insert fragment in pACC119-18, pACC119-17-0.6, pACC119-17-2 and pACC127-17-0.9, it was suggested that these clones did not cover the 5' end of the *ACC* gene.

Example 9: Cloning of the franking region of the insert fragment in pACC127-17-0.9 from the genome of *P. rhodozyma* by genome walking method

PCR primer acc26 (SEQ ID NO:8) was synthesized based on the internal sequence of pACC127-17-0.9 and used for genome walking method.

In the PCR reaction using acc26 primer, a 2.6 kb PCR band emerged from the genomic *Pvu*II library. This PCR band was cloned into pCR2.1-TOPO (Invitrogen) and it was

revealed that this clone contained 5' fragment of *ACC* gene as a result of BLAST X analysis. This clone was named as pACCPvu126 and used for further study.

Example 10: Southern blot hybridization by using pACCPvu126 as a probe

Southern blot hybridization was performed to clone a genomic fragment which covered 5' end of *ACC* gene from *P. rhodozyma*. In a similar manner as Example 7, Southern blot hybridization was performed. A probe was prepared by labeling a template DNA (*Eco*RI - digested pACCPvu116) with DIG multipriming method (Boehringer Mannheim). Hybridization was performed with the method specified by the manufacturer. As a result, a hybridized band whose size was close to 5.0 kb was visualized.

Example 11: Cloning of the genomic clone covering 5' end of *ACC* gene

In a similar manner to Example 8, the genomic fragment containing the insert fragment in pACCPvu126 was cloned by plaque hybridization. The genomic library covering 2.7 to 6.0 kb in length prepared in Example 8 was also used. Twelve positive plaques which hybridized to the insert fragment of pACCPvu126 labeled with DIG were isolated and subjected to in vivo excision to obtain plasmid DNA. As a result of sequencing for thus isolated plasmids, most of the plasmids had the identical sequence to the insert fragment of pACCPvu126. One of the clones was named as pACC204 and used for further study.

Example 12: Cloning of the gapped region between pACC204 and pACC127-17-0.9

As a result of BLAST X analysis against known acetyl-CoA carboxylase genes succeeding to the sequencing study of 3' end of the insert fragment in pACC204 and 5' end of the insert fragment in pACC127-17-0.9, it was suggested that an approximately 0.3 kb fragment could be still missing for a coverage of the entire *ACC* gene. The following PCR primers were synthesized based on the internal sequence of pACC204 and pACC127-17-0.9: acc43 (sense primer) (SEQ ID NO:9) and acc44 (antisense primer) (SEQ ID NO:10). After the PCR reaction of 25 cycles of 94°C for 15 seconds, 55°C for 30 seconds and 72°C for 15 seconds by using HF polymerase (Clontech) as a DNA polymerase and a genomic DNA obtained in Example 3 as a template, the reaction mixture was applied to agarose gel electrophoresis. One PCR band that had a desired length (0.3 kb) was recovered from the agarose gel and purified by QIAquick (QIAGEN) according to the method by the manufacturer and then cloned into pCR2.1-TOPO (Invitrogen). After transformation of competent *E. coli* TOP10, 6 white colonies were selected and plasmids were isolated with Automatic DNA isolation system. As a result of sequencing, it was found that 5 clones had an

identical sequence from each other. One of the isolated clones was designated as pACC210.

Example 13: Sequencing of a complete genomic fragment containing *ACC* gene

pACC204, pACC210, pACC127-17-0.9, pACC119-17-2, pACC119-17-0.6, pACC1224 and pACC119-18 were sequenced with primer walking procedure by using AutoRead sequencing kit (Pharmacia).

As a result of sequencing, the nucleotide sequence comprising 10561 base pairs of the genomic fragment containing the *ACC* gene from *P. rhodozyma* containing its promoter (1445 base pairs) and terminator (1030 base pairs) was determined (SEQ ID NO:1). The coding region was 8086 base pairs long and consisted of 19 exons and 18 introns. Introns were dispersed all through the coding region without 5' or 3' bias. It was found that an open reading frame (SEQ ID NO:2) consists of 2187 amino acids (SEQ ID NO:3) whose sequence is strikingly similar to the known amino acid sequence of acetyl-CoA carboxylase from other species (56.28% identity to acetyl-CoA carboxylase from *Emericella nidulans*) as a result of homology search by GENETYX-SV/RC software (Software Development Co., Ltd., Tokyo, Japan).

Fig. 1 depicts a cloned DNA fragment covering *ACC* gene region on the chromosome of *P. rhodozyma*.

Example 14: Construction of antisense plasmid for *ACC* gene

An antisense gene fragment which covers the entire structure gene for *ACC* gene is amplified by PCR and then cloned into an integration vector in which the antisense *ACC* gene is transcribed by its own *ACC* promoter in *P. rhodozyma*.

The primers include an asymmetrical recognition sequence for the restriction enzyme, *Sfi*I (GGCCNNNNNGGCC) but their asymmetrical hang-over sequence is designed to be different. This enables a directional cloning into expression vector which has the same asymmetrical sequence at their ligation sequence. The use of such a construction is disclosed in EP 1,158,051.

For the promoter and terminator fragment which can drive the transcription of the antisense *ACC* gene, the *ACC* promoter and terminator is cloned from the chromosome by using the sequence information listed in SEQ ID NO:1. The *ACC* terminator fragment is fused to a G418 resistant cassette by ligating the DNA fragment containing the *ACC* terminator to a G418 resistant cassette of pG418Sa330 (EP 1,035,206) to an appropriate vector such as pBluescriptII KS- (Stratagene).

Then, 3.1 kb of the *Sac*I fragment containing ribosomal DNA (rDNA) locus (Wery et al., Gene, 184, 89-97, 1997) is inserted downstream of the G418 cassette on thus prepared plasmid. The rDNA fragment exists in multicopies on the chromosome of eukaryote. The integration event via the rDNA fragment would result in multicopied integration onto the chromosome of the host used and this enables the overexpression of foreign genes which are harbored in expression vector.

Subsequently, *ACC* promoter is inserted in the upstream of *ACC* terminator to construct of expression vector which functions in *P. rhodozyma*.

Finally, the antisense *ACC* construct is completed by inserting the 1.5kb of *Sfi*I fragment containing antisense *ACC* into thus prepared expression vector functioning in *P. rhodozyma*. A similar plasmid construction is disclosed in EP 1,158,051.

Example 15: Transformation of *P. rhodozyma* with an *ACC*-antisense vector

The *ACC*-antisense vector thus prepared is transformed into *P. rhodozyma* wild type strain, ATCC96594. The protocol for the biolistic transformation is disclosed in EP 1,158,051.

Example 16: Characterization of antisense *ACC* recombinant of *P. rhodozyma*

Antisense *ACC* recombinant of *P. rhodozyma*, ATCC96594 is cultured in 50 ml of YPD medium in 500 ml Erlenmeyer flask at 20°C for 3 days by using their seed culture which grows in 10 ml of YPD medium in test tubes (21 mm in diameter) at 20°C for 3 days. For analysis of carotenoid produced appropriate volume of culture broth is withdrawn and used for analysis of their growth, productivity of carotenoids, especially astaxanthin. For analysis of growth, optical density at 660 nm is measured by using a UV-1200 photometer (Shimadzu Corp., Kyoto, Japan) in addition to the determination of their dried cell mass by drying up the cells derived from 1 ml of broth after microcentrifugation at 100°C for one day. For the analysis of the content of astaxanthin and total carotenoids, cells are harvested from 1.0 ml of broth after microcentrifugation and used for the extraction of the carotenoids from cells of *P. rhodozyma* by disruption with glass beads. After extraction, disrupted cells are removed by centrifugation and the resultant is analyzed for carotenoid content with HPLC. The HPLC condition used is as follows: HPLC column: Chrompack Lichrosorb si-60 (4.6 mm, 250 mm), Temperature: room temperature, Eluent: acetone / hexane (18/82) add 1 ml/L of water to eluent, Injection volume: 10 µl, Flow rate: 2.0 ml/min, Detection: UV at 450 nm. A reference sample of astaxanthin can be obtained from Hoffmann La-Roche (Basel, Switzerland).

Fig. 1 depicts a deducted biosynthetic pathway from acetyl-CoA to astaxanthin in *P. rhodozyma*.

Fig. 2 depicts a cloned DNA fragment covering *ACC* gene region on the chromosome of *P. rhodozyma*.